A Simple Python Based Inverted Index

File Structure

data.CSV
> Stores the source data to be retrieved in csv format, with each line as an entry and the first column as the title of each entry.
> Test data is already stored in this file.

sync.py
> Used to process the source data and generate an inverted-index.json file that stores the inverted index.

search.py
> Used to perform Boolean searches in inverted indexes.

docname.json
> Stores the correspondence between the document name and the number. This file will be used when displaying the search results.

inverted-index.json
> Stores the inverted index table.

Usage

First generate inverted-index.json file in command line:
```
python sync.py
```

Search in command line:
```
python search.py [your keywords to be searched]
```

Spaces between keywords will be treated as AND operators.
To use NOT or OR operators, just insert 'not' or 'or' into the keywords you want to retrieve.
For example:
```
python search.py novel
python search.py novel tana
python search.py novel or winter
python search.py not novel
python search.py not novel not computer or winter
```

NOTICE:
OR operations will always be calculated last. So '`not computer or winter`' will be treated as '`(not computer) or (winter)`' instead of '`not (computer or winter)`'.

Dependencies

    Python 3.6.5 or higher

    NLTK 3.3 or higher

Author

左振宇

1610842

计算机学院

计算机科学与技术